Large Language Models as Zero-Shot Conversational Recommenders

Abstract

In this paper, we present empirical studies on conversational recommendation tasks using representative large language models in a zero-shot setting with three primary contributions. (1) Data: To gain insights into model behavior in "in-the-wild" conversational recommendation scenarios, we construct a new dataset of recommendation-related conversations by scraping a popular discussion website. This is the largest public real-world conversational recommendation dataset to date. (2) Evaluation: On the new dataset and two existing conversational recommendation datasets, we observe that even without fine-tuning, large language models can outperform existing fine-tuned conversational recommendation models. (3) Analysis: We propose various probing tasks to investigate the mechanisms behind the remarkable performance of large language models in conversational recommendation. We analyze both the large language models' behaviors and the characteristics of the datasets, providing a holistic understanding of the models' effectiveness, limitations and suggesting directions for the design of future conversational recommenders.